The expansion of isms, 1820-1917 : Data-driven analysis of political language in digitized newspaper collections

Marjanen, Jani

2020

Marjanen , J , Kurunmäki , J A , Pivovarova , L & Zosa , E 2020 , ' The expansion of isms, 1820-1917 : Data-driven analysis of political language in digitized newspaper collections ' , Journal of Data Mining and Digital Humanities . https://doi.org/10.46298/jdmdh.6159

http://hdl.handle.net/10138/326081
https://doi.org/10.46298/jdmdh.6159

The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections

Jani Marjanen, Jussi Kurunmäki, Lidia Pivovarova, Elaine Zosa

Journal of Data Mining and Digital Humanities

The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections

Jani Marjanen[1], **Jussi Kurunmäki**[2], **Lidia Pivovarova**[1], **Elaine Zosa**[1]

[1]University of Helsinki, Finland
[2]Tampere University, Finland

Corresponding author: Jani Marjanen , jani.marjanen@helsinki.fi

Abstract

Words with the suffix -ism are reductionist terms that help us navigate complex social issues by using a simple one-word label for them. On the one hand, they are often associated with political ideologies, but on the other they are present in many other domains of language, especially culture, science, and religion. This has not always been the case. This paper studies isms in a historical record of digitized newspapers published from 1820 to 1917 in Finland to find out how the language of isms developed historically. We use diachronic word embeddings and affinity propagation clustering to trace how new isms entered the lexicon and how they relate to one another over time. We are able to show how they became more common and entered more and more domains. Still, the uses of isms as traditions for political action and thinking stand out in our analysis.

Keywords

isms; ideology; political language; diachronic word embeddings; affinity propagation clustering

INTRODUCTION

Words with the suffix -ism are indispensable terms for understanding politics and society, yet they are complex words that give rise to plenty of confusion. It is hard to tell how different isms, ranging from communism to Protestantism, and further to impressionism and positivism, really relate to one another. People using everyday language seem to uphold the link between isms, and from an analytical perspective it is clear that most words with the suffix serve some sort of reductionist function. They are words that describe something complex in just one heading [Spira, 2015].

The most common take on a particular ism is to regard it as a set of ideas that can be traced throughout history. For instance, in the case of liberalism, there is a debate over which theoreticians can be argued to have formulated key ideals of liberalism. This entails a search for the origins of an ism. A critique of such quests has started to emerge, and has shifted the focus from searching for the origins of an idea to an understanding of different historical uses of the very term. [Leonhard, 2001, Bell, 2014, Rosenblatt, 2018, Freeden et al., 2019]. The more historicizing approach has also tried to make sense of isms as a whole by producing typologies of their areas of application or characteristics [Cuttica, 2013, Höpfl, 1983].

This paper seeks to take the historical approach further by providing an analytic overview of the historical process in which new isms have emerged and developed. Isms have been used to categorize things since antiquity. In English a separate word, ism, emerged in the seventeenth century to denote them collectively. Ever since the sixteenth and seventeenth centuries, isms

have spread to many new domains in life, covering religion, politics, science, arts, and more. Isms have also gained a global reach so that they are used as cognate loans or direct translations in many languages [Höpfl, 1983, Spira, 2015, Kurunmäki and Marjanen, 2018b].

The development of isms varies depending on political context and the language used. French, German and English coinages dominate in Europe, as isms from those languages were often introduced and adapted into other European languages. Smaller languages also produced their own isms, and it is often unclear where a particular ism originated, as the easy cognate translations could be about loans across languages, but also about nearly simultaneous coinages in different places. Focusing on Finland provides a particularly interesting case in understanding these transnational developments. The Finnish data includes both Finnish-language and Swedish-language newspapers, which interacted constantly, but which also developed at different speeds. Swedish-language papers were, until the end of the nineteenth century, usually quicker to adopt new concepts from abroad, but translations to Finnish were usually quickly introduced due to many actors functioning in both languages [Marjanen et al., 2019, Engman, 2016]. In this way the Finnish case provides an example of isms being deployed in the same political context, but in one Germanic and one Fenno-Ugric language.

The Finnish case is an interesting instance of the interplay between local political contexts and different languages. We therefore cannot extrapolate results for other countries based on the Finnish case, but it is a particularly interesting point of comparison. From 1809 to 1917, Finland was a Grand Duchy in the Russian empire, and during this relatively short period of time it gained many state institutions of its own [Jussila, 2004]. As part of this process, Finnish actors also introduced new political vocabulary and developed an independent press [Hyvärinen et al., 2003]. All these processes are present in the data analyzed in this study.

By focusing on the nineteenth century and using digitized historical newspapers from Finland, this paper provides a new perspective on how isms became important in public discourse. Although linguists have paid attention to the productivity of isms [Hahn, 1981], large-scale digitized data sets provide an opportunity to look at historical language change in a statistically more robust way than before. They also allow for data-driven methods of clustering and modelling the development, which helps us chart the expansion of isms suggested by earlier research such as [Kurunmäki and Marjanen, 2018b,a]. Automatic analysis of large data collections allows us to reveal some regularities that have been hidden from researchers' attention and thus produce starting points for close reading and historical analysis. In this study, we use word embeddings to analyze the spread of isms in the Finnish context. This method, drawn from natural language processing (NLP), differs from traditional approaches in history and political science, but the ability to cluster isms in a relatively large historical data set has several benefits for scholarship in the humanities and social sciences as well. As we will show, it can partly confirm the narrative of isms becoming especially political and even ideological in the course of the nineteenth century, but also that isms relating to psychology and the sciences entered the lexicon at this time. The clustering clearly shows how these isms belonged to different language domains. Further, the method can point out interesting new findings about the scope and nature of particular isms and their use in the Finnish context, which we study semi-automatically and discuss in the results section.

I RESEARCH QUESTIONS AND DATA

1.1 Research questions

This paper studies isms as particularly laden keywords in societal discourse in Finland in the long nineteenth century. It covers a period of time when many isms were introduced into the

Swedish and Finnish languages and when the printed public sphere expanded greatly in Finland. In the early nineteenth century, only one newspaper was published in the country, whereas the number of titles had grown to around 130 newspapers by the turn of the century [Marjanen et al., 2019, Tommila and Salokangas, 1998].

We address the following research questions:

1. How did the vocabulary of isms expand over the period?
2. Which isms appear as similar based on their embeddings?
3. How does the theme of politics distinguish itself in the clusters of isms over time?
4. Are there interesting continuities in the enriched clustering that take into account the nearest neighboring words of the isms?
5. How do the language of isms in the two languages relate to one another?

The questions are partly informed by our reading of example texts in the newspapers, and some of the interpretations of research results also build on those readings, but the questions are motivated and designed to be answered primarily by computational methods.

1.2 Data

To answer our research questions, we use a digitalized collection of nineteenth-century Finnish newspapers freely available from the National Library of Finland. The collection includes every newspaper printed in Finland at the time, so changes in the size of the data follow actual publishing patterns in the country at this time [Pääkkönen et al., 2016]. Though the archive contains newspapers beginning from the 1770s, the earlier time periods do not have enough data for the analysis we apply in this paper. Thus, we keep to the data from 1820 to 1917. Even for the the period from 1820 to 1860, data is relatively scarce, particularly for Finnish, and the number of different isms is still low. Still, it is crucial to keep this period as a part of the study, as many key political isms, such as liberalism, socialism and communism, were introduced into political discourse in Europe at this time. This gives us an idea of the introduction of isms into political discourse in Finland and the interplay between the Swedish and Finnish languages. Since our data includes all newspapers published in the period, the development we trace follows the development of newspapers as a medium. Early in the nineteenth century, newspapers were usually not published more than three times a week, whereas in the early twentieth century dailies dominated the newspaper field [Marjanen et al., 2019].

The collection contains newspapers in the Swedish, Finnish, Russian, and German languages, the former two being the main languages. In our analysis, these dominant languages are treated as two separate corpora even though contemporaries often relied on newspapers in both languages. The period has been described as an interaction between three languages in Finland, Swedish being the main language of administration and learned life, Finnish being the primary language of the majority of the inhabitants in Finland and increasingly seen as the language of the future, and Russian as the language that most people in Finland did not read, but which still loomed in the background as the main language of the Russian empire [Engman, 2016].

In this paper we use the Finnish and Swedish corpora, leaving the far smaller data sets of Russian and German for further research. The total number of words in the corpora is presented in Table 1 along with the vocabulary size used to build models, that is, the number of distinct words with a count greater than 100. Both corpora are lowercased and lemmatized using LAS, an open-source language-analysis tool [Mäkelä, 2016].[1] This is a meta-analysis tool that provides a wrapper for other existing tools developed for specific tasks and languages. Though

[1] https://github.com/jiemakel/las

Table 1: Corpus and vocabulary size of the word2vec model by double decade. Corpus size in millions and vocabulary size in thousands.

Time slice	Corpus size		Vocabulary size	
	FINNISH	SWEDISH	FINNISH	SWEDISH
1820–1839	1.3M	25.5M	1.4K	14.7K
1840–1859	10.3M	77.9M	7.3K	34.7k
1860–1879	90.6M	326.7M	3.9K	99.2K
1880–1899	805.3M	966.9M	205K	227K
1900–1917	2439.0M	953.0M	534K	290K
Total	3346.6M	2355.2M	787K	666K

LAS supports multiple languages, most efforts were made to process Finnish data, including historical Finnish. The output for our Swedish data is more noisy. In particular, the Swedish LAS lemmatizer is unable to predict the lemma for out-of-vocabulary words, e.g. *boulangismen* (definite form of 'boulangism'). Thus we applied additional normalization by converting all words ending with *-ismen* or *-ismens* into *-ism* forms. For all other words we use the LAS output; implementation of proper Swedish lemmatization is beyond the scope of this paper, as most of our findings are based on clustering the isms only; thus perfect lemmatization of other words is less crucial.

II METHOD

2.1 Diachronic embeddings

To trace semantic shifts in word meanings we split a lemmatized corpus into double decades (1820–1839, 1840–1859, and so on until 1900–1917) and train continuous embeddings [Mikolov et al., 2013] on each time slice. We use the Gensim Word2Vec implementation [Řehůřek and Sojka, 2010] using the Skip-gram model, with a vector dimensionality of 100, window size 5 and a frequency threshold of 100—only lemmas that appear more than 100 times within a double decade are used for training. In this way we try to ensure that each word in a model has a reliable amount of context and that the embeddings are trustworthy.

There is no common strategy on how to choose a frequency threshold for a historical corpus with OCR errors, but earlier research does indicate of practices that mitigate problems. It has been previously observed that embedding spaces have frequency-based effects and that less frequent words have higher similarity to their neighbours, i.e. are situated closer to the center of the embedding space [Faruqui et al., 2016, Schnabel et al., 2015, Li and Wang, 2017]. This effect should be even higher for a corpus with a high number of OCR errors, since the vocabulary in this case is much larger and the contexts are sparser. Thus, we opt to cut the vocabulary at mid-frequency words to keep frequency-related problems at bay.

Apparently, we lose some isms because they appear less than 100 times in a double-decade. For example, the Finnish-language word *feminismi* was mentioned 91 times between 1900 and 1917 and was excluded from our analysis, while in Swedish its counterpart was mentioned 242 times and is visible in our results. Our models allow us to detect when a word became frequent, the context it was used in, and the difference between the two language contexts. However, they do not allow us to check when the word appeared for the first time. Furthermore, comparison of word distributions between languages is not fully reliable for less frequent words.

Since training word embeddings is a stochastic process, the particular values of vectors do not stay close across runs, though distances between words are quite stable. To ensure that

embeddings are aligned across time slices, we follow the vector initialization approach proposed in [Kim et al., 2014]: embeddings for $t+1$ time slice are initialized with vectors built on t; then training continues using new data. The learning rate value is set to the end learning rate of the previous model, to prevent models from diverging rapidly. Evaluating the quality of diachronic word embeddings is currently a challenge because of the lack of gold standard data for different languages and time periods [Shoemark et al., 2019]. We use this approach since it has previously been used in a similar study [Hengchen et al., 2019] with partly different data.

Temporally aligned embeddings have previously been used to trace semantic drift by computing the distances between vectors representing a word in two time periods or by measuring the differences in nearest neighbors for these vectors [Hamilton et al., 2016]. However, most studies that tackle semantic shift detection in computational linguistics deal with clear cases of word meaning change, such as the complete change of meaning of the word 'gay' or the acquisition of a new, completely different sense such as the words 'virus' or 'cell'. These rapid transformations could also be found in our data: e.g. the Swedish word *flygare*, which initially meant an insect but changed its meaning to "aviator" at the beginning of the twentieth century. The embedding models that we trained are able to detect this change, since the nearest neighbors of *flygare* completely changed. Distance-based methods seem to be less useful for isms, since their meanings do not change so drastically. For example, 'patriotism', whether it had positive or negative connotations, has fairly consistently had a meaning semantically close to "love of one's country". However, the political and social context in which the word was used changed over time. Further, the term could be used for quite different rhetorical purposes, and it carried new social and affective meanings that are not as readily visible in the embeddings.[2] Thus, *patriotism*, and most other isms, are vague in their meaning, making it difficult to assess what exactly is meant when they are used in historical texts. In this paper we do not depend on distances between word vectors across time to extrapolate meaning, but instead use clustering to find which isms were closer to each other—i. e., had similar contexts—over various periods of time.

There are many ways of constructing diachronic word representations other than the word embedding and alignment approach that we use here, but we opted for this method because it has been shown to produce reliable results [Schlechtweg et al., 2019] and training times are relatively short even for large corpora. Simpler methods, such as studying collocates and using them for clustering, would also require enough instances to produce reliable clusters, whereas more complex methods, such as deep contextualized embeddings or continuous time representations, have not yet been proved to produce better results for historical data with some OCR noise. For our purpose of understanding a historical development, using word embeddings is the best match for the moment.

2.2 Clustering

To investigate the expansion of the vocabulary of isms we cluster words into closed groups based on their embeddings. Since our task is mostly exploratory, and the number of clusters cannot be known in advance, we apply the Affinity Propagation clustering technique [Frey and Dueck, 2007]. This method divides all datapoints into *exemplars*, i.e., cluster representative tokens, and *instances*, i.e., other members of clusters. At the initial step, each datapoint represents a cluster on its own. Then, for each instance-representative pair a likelihood for an instance to be represented by an exemplar is computed by taking into account all other instances of the exemplar and all other available exemplars for the instance. This computation is repeated until

[2] For social, affective and other types of meaning, see Leech [1974]

convergence is reached; if an exemplar has no instances, it is dismissed. We use the standard implementation of this algorithm from the Scikit-learn package [Pedregosa et al., 2011] with default parameters.

Affinity Propagation has previously been used for various language analysis tasks, including collocation clustering into semantically related classes [Kutuzov et al., 2017] and unsupervised word sense induction [Alagić et al., 2018]. The main advantages of the method are that it detects the number of clusters automatically, and is able to produce clusters of various size. As a side effect, it returns exemplars, i.e. cluster representatives, that are not necessarily equal to the geometric centre of the cluster.

The main drawback of Affinity Propagation is pairwise computations. The method is quadratic in time and memory, and cannot be applied to large data sets, such as a whole corpus vocabulary. Thus, data selection is an unavoidable step. In this paper we use Affinity Propagation in two experiments.

In the first experiment, we extract from the corpus all ism words. i.e. words that end with *-ism* in Swedish and *-ismi* in Finnish and cluster only this set of words. We exclude from the list words that are shorter than 5 characters for Swedish and 6 characters for Finnish. This is to filter out obvious errors that appear due to OCR issues such as 'ism', 'tism', or 'rism'. Though the words 'ism' and 'ismi' exist in the Swedish and Finnish languages, they are very uncommon in nineteenth-century press. The extraction allows us to identify how close these words are to each other given other isms in the corpus.

In the second experiment, we try to put isms into a richer context and trace other words associated with them in the respective double-decades. We extract from the corpus all words that have a cosine similarity of less than 0.5 to any isms. Then we perform clustering on this enriched data set. Finally, the clusters are filtered so that only clusters that contain at least one isms word are presented for qualitative analysis. The output of this procedure is different from that of the first experiment, i.e. words that were clustered together in the isms-only clustering can break up into different enriched clusters, since in the latter setting they have more exemplar options.

Henceforth we refer to the results of the first and the second experiments as *ism clusters* and *enriched clusters* respectively. We discuss the outcomes of the two experiments alternately since they provide different perspectives on the development of ism vocabulary. The first experiment helps us to understand the main question about the expansion of isms, whereas the the second experiment provides additional results for interpretation and is used especially in the section on separatism.

Clustering is performed separately for each time slice. To link clusters across time, we perform visualization with Sankey charts. In the Sankey diagram, clusters from time slice t are linked to clusters in time slice $t+1$ if they have words in common. The magnitude of the link is the sum of the word frequencies of the common words between the linked clusters from adjacent time slices. We use the frequencies from the source cluster, that is, the cluster from time slice t.[3] This is not the only way to visualize the links. We also tried visualizing the number of shared words in the cluster, but the visualization based on frequency provides a clearer picture of the development. Using the number of shared word might work better if the clusters include many infrequent words, but in this case the frequency threshold set for training the model excludes low-frequency words.

[3] Code for our experiments is available at https://github.com/ezosa/Diachronic-Embeddings

III RESULTS

Some of our results are directly related to the political history of Finland and the development of newspapers as a medium, whereas others go well with previous notions of the development of the language of isms in general. They strengthen earlier interpretations by providing more robust proof for interpretations that have mostly relied on the qualitative reading of sources. Other findings are surprising for historians of political ideologies, and may compel us to rethink how we see the history of political discourse. In what follows, we will present the findings in this order.

3.1 Swedish-language and Finnish-language clusters in comparison

As expected, Finnish-language and Swedish-language isms cluster differently in terms of timing and themes present (see Figure 3 and Figure 4). There are three main reasons for this:

1. The Swedish-language press in Finland developed earlier and included more abstract content earlier in the century, whereas newspapers in Finnish—and the Finnish written language—matured only in the latter half of the century. Consequently, we have been able to produce meaningful clusters of isms for 1820s onward for Swedish and only from the 1860s onward for Finnish. As described earlier, the languages were in constant interaction, but the scope of Finnish-language newspapers was much smaller in the first half of the century and their content was less theoretical and political [Nurmio, 1934, Rantala et al., 2019]. Furthermore, Swedish-language newspapers were quicker to adopt new terms from publications in Sweden because of the language connection, and thus performed a mediating function with regard to new political vocabulary [Zilliacus and Knif, 1985].
2. The *-ismi* was not a productive suffix in the Finnish language, but was used through cognate loans and through analogous derivation of foreign words.[4] Consequently, isms are generally less common in Finnish than in Swedish. Nonetheless, they were used in both languages, especially as Finnish political language developed through an interplay with Swedish [Stenius, 2004]. In the particular case of adopting isms as key terminology in Finnish, the latter half of the century was a crucial turning point.
3. The political outlook of the two languages was slightly different. From the 1880s onward, the Finnish and Swedish newspapers were printed in nearly equal quantities. At this time, the language spheres also started specializing. Swedish speakers lived mostly in larger towns and around the coast, whereas Finnish speakers inhabited most of the country [Marjanen et al., 2019]. In Lapland, Sami languages also had a strong presence, but they did not appear in print at this time. At this point, Finnish-language papers were more likely to have a rural or working-class background and Swedish-language papers were more likely to be more urban, liberal and bourgeois [Engman, 2016], which also shows in the use of isms. This is typically visible in the proportionately greater role that the cluster around socialism manifests in Finnish compared to Swedish. The clusters clearly show that Finnish-language ism vocabulary was more politically oriented in the early twentieth century. Cultural, philosophical and scientific isms were less present.

The distinction between Swedish and Finnish is also visible from the analysis of the enriched clusters. The number of words used in various steps of analysis is presented in Table 2, which shows that the number of isms in the Finnish data is much smaller than for the Swedish data. The table also shows that although 0.5 is an arbitrary threshold, up to 90% of words selected using

[4] The ism is not strictly speaking a suffix in Finnish, but a rather a sublexical suffix-like unit as often the entire words are cognate loans in which the root itself is not a word in Finnish. We thank Antti Kanner for pointing this out.

FINNISH

Time slice	***ism***	***close***	***cluster***	***select***
1820 - 1839	0	-	-	-
1840 - 1859	0	-	-	-
1860 - 1879	1	157	1	12
1880 - 1899	35	5977	20	442
1900 - 1917	119	8940	70	1543

SWEDISH

Time slice	***ism***	***close***	***cluster***	***select***
1820 - 1839	3	724	3	49
1840 - 1859	17	1845	12	211
1860 - 1879	61	5229	31	669
1880 - 1899	120	12233	54	1320
1900 - 1917	137	11858	56	1387

Table 2: Number of distinct words used in various steps of the process to obtain enriched clusters: **isms** is the number of distinct words with suffix *-ism*, ***close*** is the number of words that have a cosine similarity of higher than 0.5 to at least one ism, ***cluster*** is the number of clusters that contain at least one ism, ***select*** is the number of words in these clusters.

this threshold are filtered out after the clustering, i.e. they fall in clusters that do not contain any ism word. Ism words, on the contrary, are not spread across clusters but concentrated in only a few of them. As can be seen from the table, the number of selected clusters is generally smaller than the number of words with the suffix ism since they tend to cluster together. This is an indirect justification that the threshold is sufficient and most of the relevant words are present in the output, since the majority of the candidate words are filtered as irrelevant.

3.2 Expansion of the language of isms

By looking at the relative frequency of different isms over time, we see an expansion of isms in the nineteenth century (see Figure 1). This is partly the function of a growth in data size over time, but mostly because new isms were introduced and often also lexicalized to the extent that they became nodal points in newspaper discourse. Isms like socialism and communism entered the lexicon in the 1830s and 1840s in many European languages, and are almost simultaneously visible in the Finnish materials. A similar pattern is visible with other part of human activity, with words such as spiritism or modernism being introduced in the latter half of the nineteenth century. New political, social and cultural phenomena were categorized through new isms and the notion of isms itself expanded.[Kurunmäki and Marjanen, 2018a]

While some individual isms became very common and grew in frequency, this is not the case for all of them. Some stagnated and others were simply short-lived coinages. What matters is the overall productivity of isms that is visible in the unique number of isms used in the newspapers per year (see Figure 2). The overall growing trend in relative frequency corresponds with similar developments in English, as evidenced in the Google Books data set. [Kurunmäki and Marjanen, 2018a]

One feature of the suffix is that it is easy to deploy in *ad hoc* inventions of new words, which means that many isms were introduced but never resonated in public use. These are interesting instances of linguistic innovation as such, but are excluded from this study as we use a frequency threshold for training our embeddings. The threshold also effectively excludes many false variants caused by noisy optical character recognition.

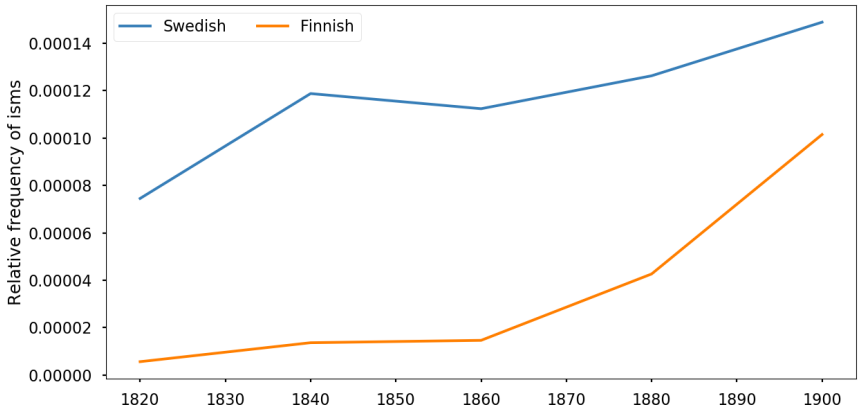


Figure 1: Relative frequency of all ism words found in the Finnish and Swedish corpora by time slice.

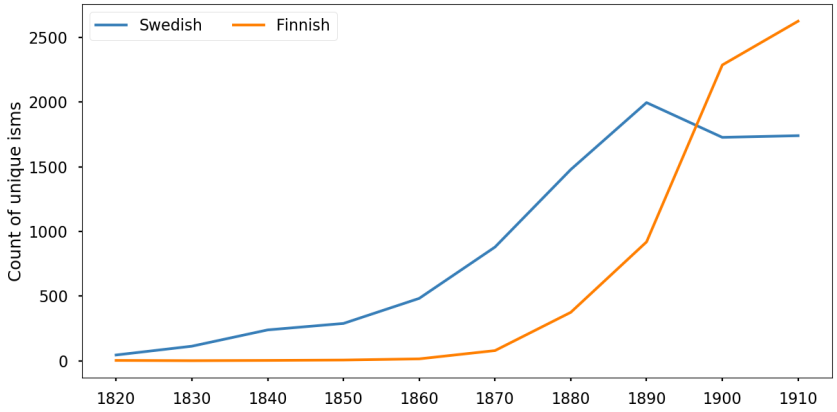


Figure 2: Counts of distinct ism words in the Finnish and Swedish corpora by time slice. This includes words that have OCR issues and thus do not appear frequently in the corpora.

Aligning the clusters in the Sankey plots makes it possible to visually explore how the vocabulary of isms developed over the course of the century. As can be seen in Figure 3, there is a steady expansion of isms from the 1820s onward for Swedish. As the models for producing the clusters rely on enough datapoints for training, particular clusters appear with a delay compared to first uses of particular words. For instance, *patriotism* appears in the corpus for the first time in 1791 and *liberalism* 1820, but the clusters of which they are part (but not necessarily cluster representatives or most frequent ones) appear in 1820–1839 and 1840–1859 respectively, as can be seen in Swedish clusters (Table 8). The word *socialism* appears the first time in 1840 and is also included in the cluster for 1840–1859, since it immediately became popular and the number of newspapers in Swedish had already grown.

The visualization of Finnish-language clusters (Figure 4) provides a much shorter story, but the expansion of isms into new domains is also visible in this data. Compared to the Swedish clusters, it is remarkable that the Finnish language of isms began with socialism, which then, so to say, invited other isms that had been available for quite some time in Swedish. This is explained by three different factors. First, the -ism was not a standard suffix in Finno-Ugric languages at this time, so there was a reluctance to using ism words in cognate translations. Second,

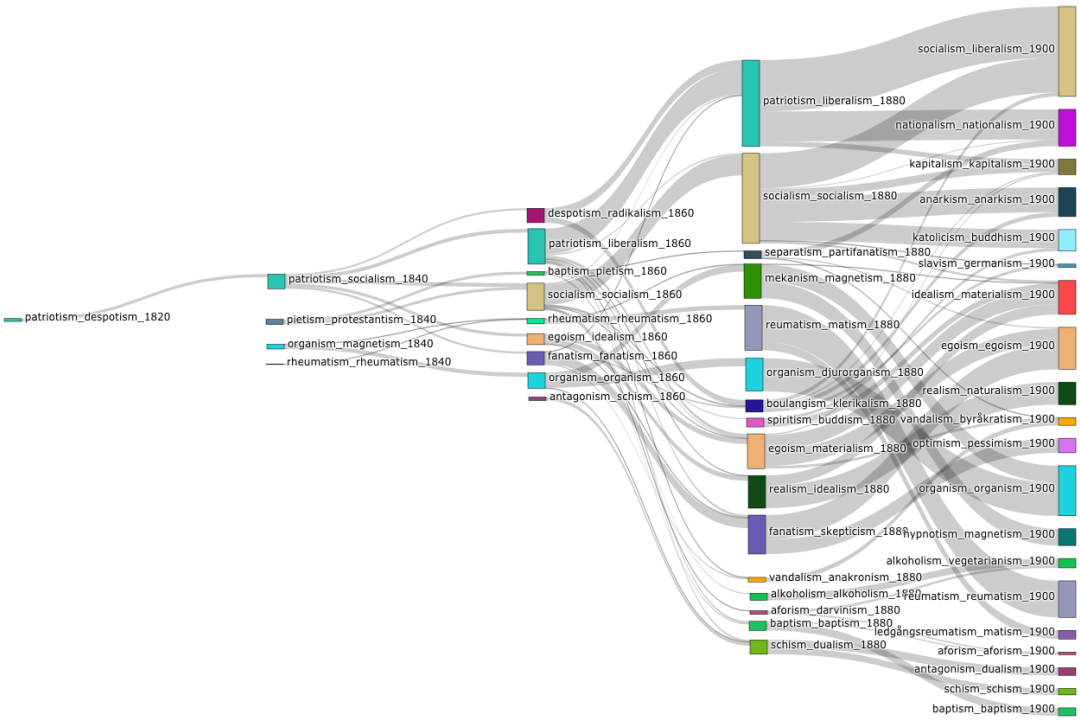


Figure 3: Sankey diagram of isms clusters from the Swedish data set covering five double decades from 1820 to 1917. A cluster name is the most frequent ism word for that cluster, followed by the cluster representative and the double decade. Cluster size is the sum of the cluster word frequencies. Band width shows the weighted proportion of common words.

Finnish-language newspaper publicity really started growing after 1860 [Marjanen et al., 2019]. Third, in the 1850s, Finnish-language publications were censored more severely than Swedish-language publications [Nurmio, 1947], which meant that they there was less of a tradition of writing about political issues in Finnish.

Another peculiarity of the Finnish data is that although the number of isms also grew for Finnish, the clusters show much stronger continuities. The clusters fluctuate much less than for the Swedish data and the clusters that have socialism as either the most frequent word in the cluster or the cluster representative share few words with other clusters. The Finnish case is dominated by clusters that can be described as political or ideological, and these also changed the landscape of isms, whereas medical, cultural and scholarly isms played only a minor role. In this sense the language of isms comes across as much more focused and much more consistent than in Swedish (and probably also other Germanic languages). While most ism words in Swedish do have cognate translations in Finnish, it would be reasonable to interpret that the discourse was similar in both languages, but our way of clustering the use of isms according to frequency in the totality of newspapers in this period also points at this clear difference in the actual discourse as a whole.

3.3 Politics and ideology as distinct clusters

Previous interpretations by Kurunmäki and Marjanen [2018a], have suggested that the early nineteenth century meant the breakthrough of isms that we associate today with major political ideologies, whereas the end of the century saw the rise of plenty of new isms in the sciences

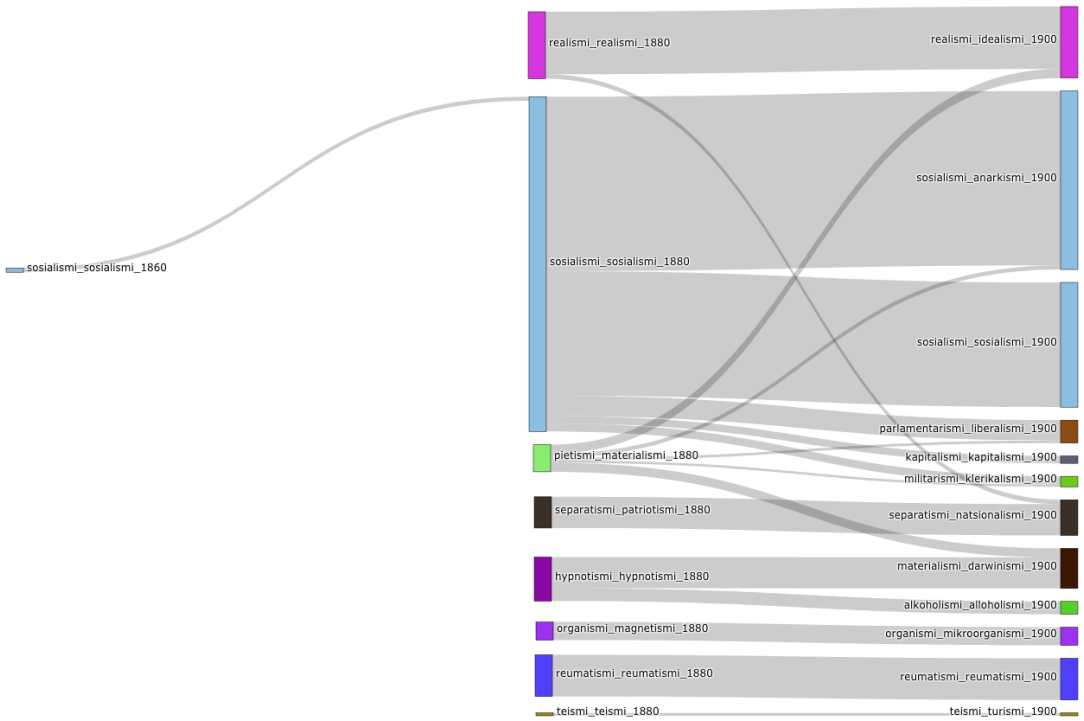


Figure 4: Sankey diagram of isms clusters from the Finnish data set covering three double decades from 1860 to 1917. The cluster name is the most frequent ism word for that cluster, followed by the cluster representative in the double decade.

(including medicine) and the arts. Again, looking at first appearances of particular isms in the Swedish-language data set suggests that this also holds for Finland. However, the clusters allow for a stronger claim, suggesting that the political and ideological isms formed a distinct category after they were introduced. This is even more pronounced for the Finnish-language data.

The clustering results, presented in Figure 3 and Table 8, allow us to trace more political and ideological clusters. A cluster size in the Sankey diagram corresponds to the sum of the word frequencies in a given double decade, while the width of a band connecting two clusters shows the proportion of cluster words (weighted by frequency) shared between these clusters. The table contains a complete list of cluster words and their frequencies for all periods.

As can be seen from Figure 3 and Table 8, there is a clear continuity in the politically laden isms that start from a cluster with *patriotism*, *fanatism* (Eng. fanaticism) and *despotism* in one cluster in 1820–1839 and continue to expand over the succeeding double decades. The most frequently occurring isms in the political clusters are *patriotism*, *socialism* and *despotism* up to 1859, and then *boulangism*, *fanatism*, *anarkism* (Eng. anarchism), *nationalism* and *kapitalism* (Eng. capitalism) up to 1917. There is some fluctuation between the political clusters, with *liberalism* and *patriotism* being quite tightly associated with one another until the last time slice of the investigated period, and some unsurprising continuities, like *konservatism* (Eng. conservatism) and *liberalism* being in the same clusters throughout. Still, it seems that there is less fluctuation between the clusters we call political and the clusters with a different focus. Religious isms (starting from *pietism*), and medical isms (e.g. *rheumatism*) come across as fairly stable. Philosophical, artistic and scientific isms are also distinguishable, albeit they do cluster

more freely. The case of rheumatism is very specific as it has a high frequency and appears often in health-related advertisements, which means that it does not co-occur very often with other isms, but is instead an isolated term in the marketing of pills and ointments.

For Finnish-language material, the data is too scarce to produce meaningful clusters for more than three time slices, as is clear from Figure 4. Even though the Finnish corpus for the 1880–1899 double decade is comparable in size with the Swedish corpus, the number of *distinct isms* in Finnish is smaller than in Swedish: 44 for Finnish and 125 for Swedish.

With scarcer data the distinctness of the clusters is even clearer. Clusters with *socialism* as the most frequent ism are dominant both for Swedish and Finnish, but the role of *socialism* as a pivotal ism is even more pronounced for the latter, as is also indicated by Marzec and Turunen [2018].

3.4 Socialism as a pivotal ism

While the two data sets are different, they both show that many isms pivot around the discourse of socialism, especially toward the end of the century. *Socialism* does not fluctuate between clusters, but really seems to be one of the terms that organized the debate. We obtain a supplementary perspective on this phenomenon by looking at the relative frequency of a selection of the most frequently occurring isms in our data (Figure 5). Like the clusters, the relative frequencies indicate a growing proportion of isms over time and also reveal some differences between the data sets. For the Swedish data set, we see a change in the overall landscape of the vocabulary, with terms such as *patriotism* being dominant at first but then surpassed in frequency by *socialism*. In Swedish, we also find a broader selection of isms from political to religious and medical topics, present for the second half of the nineteenth century.

In Finnish, the landscape is different as it appears that the whole vocabulary relating to isms was dominated by *socialism* from the 1860s onward. It seems that the word *socialism* paved the way for other isms to be lexicalized in the Finnish language. Once socialism became ubiquitous in Finnish-language political discourse, other isms well-known from Swedish and other Germanic languages were easier to introduce to Finnish. This does not mean that isms had not featured in Finnish at all, only that they had been infrequent and not a normal part of the lexicon. We must also note that most authors who produced texts in Finnish also operated in Swedish, so while they did not write about isms in Finnish, they still held notions of isms through the other main language of the country.

Figure 5 also shows that *capitalism* was an ism that became more commonly used in the early twentieth century. This follows international trends, but in this case it is perhaps most interesting to note that the use of *capitalism* is dominant in socialist newspapers – even more so than for the word *socialism*. We can see this clearly if we look at the titles in which the word occurred. For instance, a random sample of 1,000 occurrences of the word *capitalism* in 1907 is distributed over 87 different newspapers, most of which display the word only 1–4 times that year. The papers that most commonly used the term had a distinct, socialist or social democratic profile. The papers *Työmies* (The worker, 164), *Työ* (Work, 65), *Kansan Tahto* (The People's Will, 54), *Savon Työmies* (The Savonian Worker, 52), *Kansan Lehti* (The People's Newspaper, 52), *Sosialisti* (The Socialist, 47), *Sorretun Voima* (The Power of the Oppressed, 45), *Elämä* (Life, 42) together account for more than half of the hits. The first non-socialist newspaper with 13 occurrences was the leading Fennoman paper *Uusi Suometar*. All in all, socialist or social democratic papers accounted for more than 800 of the uses of capitalism in the papers.[5]

[5] The figures were pulled from http://korp.csc.fi by loading concordances for the word *capitalism*

Table 3: Enriched clusters for Finnish and Swedish that contain the word *socialism*(*i*). Cluster *representatives* are marked in italics, **isms** are bolded.

1880–1889			
FINNISH		SWEDISH	
sosialismi 'socialism'	5115	**socialism 'socialism'**	5560
***anarkismi* 'anarchism'**	1120	*reaktion* 'reaction'	6991
***nihilismi* 'nihilism'**	602	*socialdemokrati* 'social democracy'	2303
***militarismi* 'militarism'**	328	***anarkism* 'anarchism'**	1975
***kommunismi* 'communism'**	316	*frigörelse* 'liberation'	1823
***radikalismi* 'radicalism'**	171	*proletariat* 'proletariat'	1548
sosiaalidemokratia 'social democracy'	386	*emancipation* 'emancipation'	1225
sosialidemokratia 'social democracy'	339	***nihilism* 'nihilism'**	1181
villitys 'craze'	337	*socialdemokratien* 'social democracy'	1023
luokkataistelu 'class struggle'	177	*utopi* 'utopia'	1016
reaktio 'reaction'	136	***antisemitism* 'antisemitism'**	911
pappis-malta 'clericalism'$_{ocr}$	130	*bourgeoisie* 'bourgeoisie'	772
		anti 'anti-'$_{ocr}$	747
		elementerna 'elements'	703
		***absolutism* 'absolutism'**	641
		***klerikalism* 'clericalism'**	569
		***statssocialism* 'state socialism'**	485
		***kommunism* 'communism'**	459
		***ateism* 'atheism'**	455
		kvinnoemancipation 'women's emancipation'	445
		***panslavism* 'panslavsim'**	341
		reaktionen 'reaction'	335
		kvinnorörelse 'women's movement'	332
		framtidsstat 'future state'	242
		***kapitalism* 'capitalism'**	226
		***jesuitism* 'jesuitism'**	206
		***individualism* 'individualism'**	196
		socia 'social'$_{ocr}$	174
		ateistisk 'atheistic'	173
		fredsidé 'idea of peace'$_{ocr}$	155
		***ultramontanism* 'ultramontanism'**	129
		***utilitarism* 'utilitarianism'**	124
		kollektivistisk 'collectivistic'	122
		***kollektivism* 'collectivism'**	121
		***cesarism* 'cesarism'**	110
		frihetsidé 'idea of liberty'	108

It is clear that the increasing levels of discourse around *capitalism* were related to the rise of socialist newspapers and their political rhetoric. It was not uncommon to read about the "shackles of capitalism" or other very negatively laden statements in this discourse[6], and with this rhetoric it was unlikely for bourgeois papers to deploy the same vocabulary. If *capitalism* appeared more often in socialist newspapers, this is true also for *socialism* (although to a lesser degree), as the term became a term of self-identification for most (but not all) left-wing newspapers. Echoing one of the most famous slogans of the Fennomans, "We are not Swedes, we do not want to become Russians, so let us be Finns", the newspaper *Elämä* reshaped this pseudo syllogism as

and then counted according to newspaper titles.

[6] See e.g. *Kansan Lehti*, 24 October 1903, p. 2

Table 4: Enriched clusters for Finnish and Swedish that contain the word *socialism*(*i*). Continuation.

1900–1917			
FINNISH		SWEDISH	
***sosialismi* 'socialism'**	75117	**socialism 'socialism'**	15080
kristitty 'christian'	72175	*socialdemokrati* 'social democracy'	11030
kristinusko 'christianity'	32542	*klasskamp* 'class struggle'	2998
kristillisyys 'christian'	18566	***anarkism* 'anarchism'**	1709
rauhanaate 'pacifism'	1598	*socialdemokratien* 'social democracy'	993
***kommunismi* 'communism'**	1548	***absolutism* 'absolutism'**	879
pakanakansa 'pagan people'	760	*framtidsstat* 'future state'	533
buddhalaisuus 'buddhism'	456	***individualism* 'individualism'**	512
lristinuslo 'christianity'$_{ocr}$	428	*demokratien* 'democracy'	496
tinusko 'christianity'$_{ocr}$	383	***skandinavism* 'skandinavism'**	440
käännytys 'conversion'	256	***syndikalism* 'syndicalism'**	387
tristi '?'$_{ocr}$	252	*fredstank* 'pacifism'	342
adventisti 'adventist'	243	***antisemitism* 'antisemitism'**	341
alliansi 'alliance'	164	***marxism* 'marxism'**	286
kristinuslo 'christianity'$_{ocr}$	161	***internationalism* 'internationalism'**	285
tinuslo 'christianity'$_{ocr}$	147	***antimilitarism* 'antimilitarism'**	267
buddalaisuu 'buddhism'$_{ocr}$	144	***kommunism* 'communism'**	256
tristinusko 'christianity'$_{ocr}$	128	*historieuppfattning* 'understanding of history'	236
jumalausko 'faith'	127	*studentrörelse* 'student movement'	170
islami 'islam'	123	***aktivism* 'activism'**	168
buddalaisuusi 'buddhism'$_{ocr}$	119	***revisionism* 'revisionism'**	166
konfusius 'confucius'	118	*brandfackla* 'bombshell'	142
lristinusko 'christianity'$_{ocr}$	114	*kulturrörelse* 'cultural movement'	134
järkeisoppi 'philosophy'	111	*förbudsrörelse* 'prohibition movement'	122
tristinuslo 'christianity'	109	***försvarsnihilism* 'defence nihilism'**	117
alkukristillisyys 'early christianity'	103	***nykterism* 'prohibition movement'**	112
		***ungsocialism* 'ungsocialism'**	112
		***kollektivism* 'collectivism'**	110
		***modernism* 'modernism'**	109
		samhällsrörelse 'social movement'	102
		finskhetsrörelsen 'finnish movement'	101

a rallying call for socialists: "We are not capitalists, we do not want to become anarchists, we must then become socialists".[7]

The rapid breakthrough of isms in the Finnish-language material, in contrast to that in Swedish in Finland and Western Europe in general, testifies to an abrupt change in political language. Once socialism had been introduced into Finnish, other isms followed very quickly. We see this as a synchronization of Finnish and Germanic political thought, so that ideologically laden words with the suffix -ism were introduced as cognate translations and functioned as a way of placing Finnish political discourse on par with that in Swedish in the same country, as well as other Germanic languages in Europe [Jordheim, 2014, 2017]. This naturally also holds for non-ideological isms, but the point is especially important for comparing ideological positions.

Our findings about *socialism* as a pivotal ism in both Swedish- and Finnish-language discourse in Finland harmonize with Marzec and Turunen [2018], who emphasize the role of *socialism* based on frequency and textual analysis, but we further note that looking at *socialism* in

[7] *Elämä*, 14 March 1907, p. 1. For a history of the Fennoman slogan, see [Marjanen, 2020].

the context of all isms shows that it also had a synchronizing function between Finnish and Swedish. The breakthrough of *socialism* as a buzz word in the second half of the nineteenth century helped produce political and ideological isms in Finnish that could be compared with counterparts in Swedish and other languages.

A careful analysis of text would provide more reliable interpretations as to why socialism gained such a dominant role in Finnish-language discourse, but our enriched clustering with a cosine similarity to any word also provides more information about the linguistic contexts of each ism. Tables 3 and 4 show how Finnish-language clusters with words associated with socialism include more religious (and to certain extent also scientific) terminology than the more political discourse visible in the Swedish-language clusters. In Finnish, this development is especially clear going from the period 1880–1899 to the period 1900–1917, in which *socialism* clusters with words like "Christian" (person), "Christianity", "Christian" (adjective), "pacifism", "communism", "pagan" (person), and "Buddhism". For the Swedish clusters this shift does not take place; the cluster remains couched in the world of ideologies, future visions, politics.

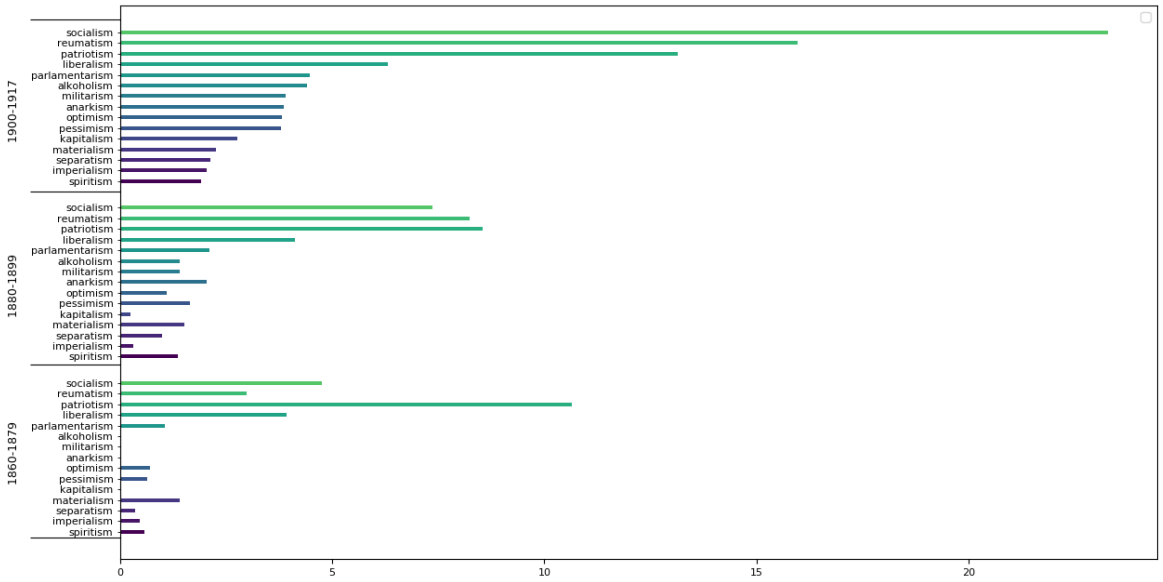
Why socialist discourse was more prone to tap into a reservoir of religious rhetoric in Finnish than in Swedish requires further study. One possible explanation may lie in the fact that socialism was related in Finnish to a higher degree than in Swedish to the so-called social question, that is, the political problematization of class, poverty and labor issues. These issues also dovetailed with Finnish-language religious discourse around the turn of the century [Alapuro et al., 1987]. Examples in the newspapers show that socialist outlets often used religious rhetoric because it was a genre that people were familiar with [Marzec and Turunen, 2018, Kemppainen, 2020]. The newspaper *Työkansa* even outlined the principles of "Christian socialism".[8], a theme that was far from marginal in the period. Another dimension of the link between socialism and Christianity comes from more theoretical discussions about the relationship between the two. At this time it was common to juxtapose socialism and Christianity, and many texts reacted to this by either trying to show their incompatibility or show that this claimed contradiction was false. Some texts took a historical perspective on this question, as was the case with the intellectual magazine *Vartija*, which spent much space on exploring early Christianity as a model for communism and socialism. [9]

3.5 Separatism and its different domains

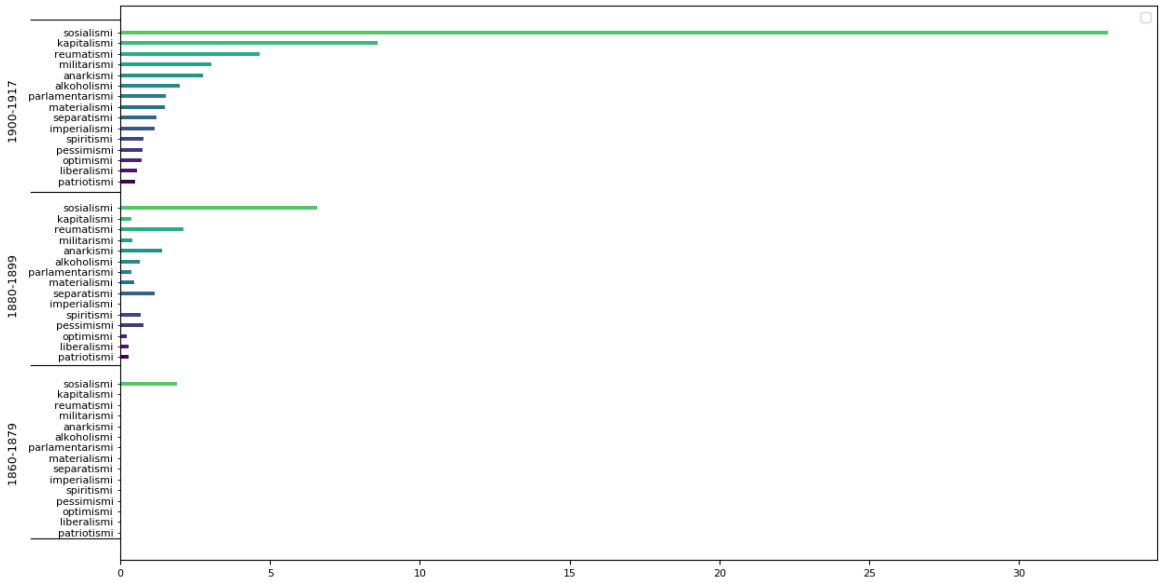
If words like *socialism* and *rheumatism* show remarkable continuity through clusters, other isms seem to be less tied to their clusters. A surprising and illuminating example of this is *separatism* in both Swedish and Finnish. In Table 5, we present the enhanced clusters for it in the Swedish data set.

Most of the words similar to *separatism* in the 1860–1879 cluster are religious, philosophical or scientific notions, such as mysticism, Darwinism, human nature, negation or idealistic. By analyzing the clusters and reading sample texts from the period, we conclude that the cluster derives strongly from debates about religion and the historical experience of Lutheranism being threatened. The paper *Vasabladet*, for instance, wrote about Evangelical movements as embodiments of "sectarian character and separatism from the church".[10] In the period, new scientific and philosophical strands of thought as well as contemporary religious revival movements seriously challenged the status of the dominant state church in Finland. The notion of separatism seems to have been used often in the ensuing debates.

[8] *Työkansa*, 2 December 1907, p. 2.
[9] See e.g. *Vartija*, 1 June 1907, pp. 195–203.
[10] *Vasabladet*, 26 September 1877, p. 1



SWEDISH



FINNISH

Figure 5: A selection of the most frequent words ending with suffix *-ism*/*ismi*. The x-axis presents relative frequency in items per million.

The 1880–1899 cluster contains a completely different set of words, including references to ethnicity and language policy in the country, such as Finnishness, Fennomans, and language policy, and contains emotional expressions such as *agitation* and *fanaticism*. The outlier of *photophobia* (*ljusskygghet*) also belongs to a similar discourse, as the term was used metaphorically at the time to discuss things that could not be brought into light because of political tensions. Again with selected reading of texts we note that *separatism* is clearly clustered with words that are related to a contemporary discussion about national identity and national language in Finland, but also more broadly within the Russian Empire. Many of the texts actually reported

on news in Russian newspapers, as in the case of the paper *Finland*, which wrote of how the "Slavophile Russian press is in a continuous state of nervousness, in which it everywhere sees opponents to the Russian idea of state. First one corner of the country, then another, is accused of separatism."[11]

The 1900–1917 cluster is different from the previous two and contains more general political lexis. Again, it seems that the notion of separatism had been included in a new discursive domain. Now, the word *separatism* clusters with words that relate to state structures and even the context of the Russian empire. Separatism had become embedded in discussions about independence, the role of Finland and as a nation. There is some continuation from the previous double decade, especially with regard to Finland's position in the Russian empire, but it still seems that the discourse on separatism shifted focus. For instance, the paper *Wiborgs Nyheter* wrote in 1913 about how "revolutionary separatism in Finland had not reached all layers of society".[12]

All in all, in three consecutive double decades *separatism* had a mostly religious context at first, but was soon adopted into a discourse relating to ethnicity and the language question, which was central to the period, and finally it spread into a more general political discourse in which separatism was more abstract. There is a certain continuity throughout the time periods, and the latter two phases are clearly related to one another. Here, the reading of individual articles and analysis of the changing enriched clusters complement each other. The former highlights continuities, whereas the latter points at the differences by bringing out the dominant words that cluster with *separatism* in the different time slices.

The Finnish-language clusters for *separatismi*, presented in Table 6, suggest a similar development, but given the political struggle regarding language preferences in the country, the perspective is slightly different. The Finnish data set does not include a cluster for the period 1860–1879 as the word occurs less than a hundred times and is therefore excluded from our models. The periods for 1880–1899 and 1900–1917 point at separatism being first dominated by the language question and then in the early twentieth century being dominated by the issue of Finland's status in the empire, and nationalism in general. Interestingly, however, the Finnish-language cluster for 1880–1899 contains more words that relate to the Svekomans, that is the Swedish-language movement. The cluster includes words like "Svekoman", "Swedish-minded", and "Viking", doing so in several different spellings and variations. This juxtaposes with the Swedish-language cluster, which includes more words relating to Fennomans as well as terms that relate in general to the tensions between the language groups. Together with a reading of a selection of the sources, we see how the discourse on separatism revolved greatly around the language question and was often about blaming the "opposing" side. However, the historical situation was more complex, as identification with language did not necessarily go hand in hand with class identification, political views or relation to the Russian empire. This holds especially in the period around the breakthrough of universal suffrage 1906 [Kurunmäki and Liikanen, 2018]. Often the rhetoric of separatism was evoked on the level of rumors and fears relating to Russia, as was the case of *Uusi Savo* reporting on the Swedish Party's fear that their political action would be labelled by the Finnish party as an act of separatism.[13]

The Finnish-language cluster for the period 1900–1917 is also similar to the Swedish-language counterpart in emphasizing the imperial context. It also seems that the different languages seem

[11] *Finland*, 5 Feburary 1885, p. 3.
[12] *Wiborgs Nyheter*, 20 November 1913, p. 3.
[13] *Uusi Savo*, 5 October 1893, p. 2

to converge in their outlook, as the question of language identity was no longer as topical in this particular discourse in Finnish or in Swedish.

The change in the distribution of *separatism* seems to be related to a change in the dominant context in which it was discussed (from a religious context to a political context). The shift in cluster entails some degree of semantic change, but it is also clear that *separatism*, as a highly abstract term, could lend itself to many different themes or topics, and thus it seems that the change in dominant themes themselves is more important for the changing clusters than the changes in the meaning of the word. An alternative interpretation would be that separatism was a polysemous word in which the different separatisms (those relating to religion, the language issue or the national question) coincided, and that different senses dominated in different time slices, but a reading of sample sentences does not support this interpretation.

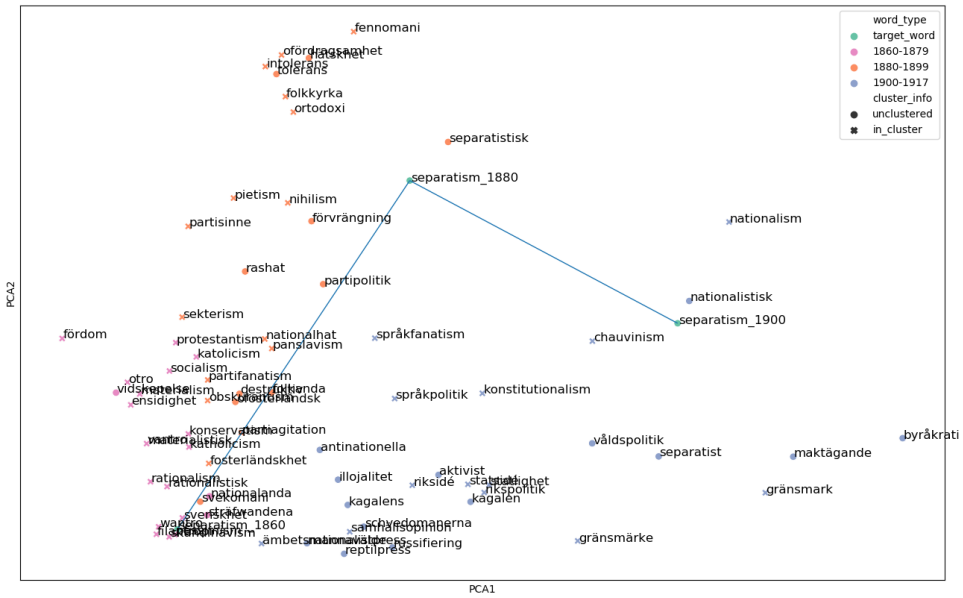
The distributional shift of *separatism* is to some extent visible from changes in the nearest neighbors of the word presented in Figure 6. They visualize a shift from the time slice 1880–1899 to 1900–1917 in both languages. The outlook can be interpreted in a similar way as the clusters produced by Affinity Propagation, but has a slightly different selection of words. This can be explained by the nature of the procedure used to produce the visualization. PCA is a dimensionality reduction technique and does not explicitly perform any clustering. Therefore each word can be among the nearest neighbors for any number of other words while Affinity Propagation assigns a word to exactly one cluster so that, for instance, *socialism* and *katolicism* are separated in clusters of their own. The difference between outputs demonstrates an added value of the clustering, which selects only one word split among many possibilities provided by embeddings. At the same time, this also means loss of information, especially for polysemous words.

1860-1879	1880-1899	1900-1917
separatism ‘ separatism’	*separatism* ‘ separatism’ *rent* ‘?’	*separatism* ‘ separatism’ *riksidé* ‘ national idea’$_{ocr}$
mysticism ‘ mysticism’ *naturalism* ‘ naturalism’	*finskhet* ‘ Finnishness’ *fennomanins* ‘ Fennomania’	*statsidé* ‘ state idea’$_{ocr}$ *rikspolitik* ‘ national policy’
darwinism ‘ darwinism’ *moral* ‘ morality’	*fennomani* ‘ Fennomania’ *svenskhet* ‘ Swedishness’	*bourgeoisins* ‘ bourgeoisie’ *byråkratien* ‘ bureaucracy’
tidsanda ‘ zeitgeist’ *krass* ‘ crass’ *utopi* ‘ utopia’	*fennomanin* ‘ Fennomania’ *vikingaparti* ‘ Viking party’	*samhällsopinion* ‘ societal opinion’
materialistisk ‘ materialistic’ *otro* ‘ incredible’	*språkpolitik* ‘ language policy’ *publicistisk* ‘ journalistic’	*sträfvandenas* ‘aspirations’ *rikskomplex* ‘ national complex’
rationalistisk ‘ rationalistic’ *wantro* ‘misbelief’	*partiagitation* ‘ party agitation’ *partiyra* ‘party delirium’	*nationalitet-* ‘ nationality’$_{ocr}$ *santryska* ‘ true Russian’$_{ocr}$
menniskonaturen ‘ human nature’ *tidehvarfvets* ‘the age (genitive)’	*partifanatism* ‘ party fanaticism’	*ämbetsmannavälde* ‘ officialdom’
materialism ‘ materialism’ *materialist* ‘ materialistic’	*språkgräl* ‘ language quarrel’	*gränsmärke* ‘ borderline’ *gränsmark* ‘ borderline’$_{ocr}$
konservatism ‘ conservatism’	*språkfanatism* ‘ language fanaticism’	*riksenhet* ‘ national assembly’
idealism ‘ idealism’ *rationalism* ‘ rationalism’	*språkfråga* ‘ language question’	*samhällskraft* ‘ social force’ *statlighet* ‘ statehood’
negation ‘ negation’ *abstraktion* ‘ abstraction’	*språkfrågan* ‘language question’	*frihetssträvande* ‘freedom-aspiring’ *väldets* ‘domination/empire’
idealistisk ‘idealistic’	*ljusskygghet* ‘ photophobia’	*riksmakt* ‘ national power’ *självhärskarmakten* ‘autocratic power’

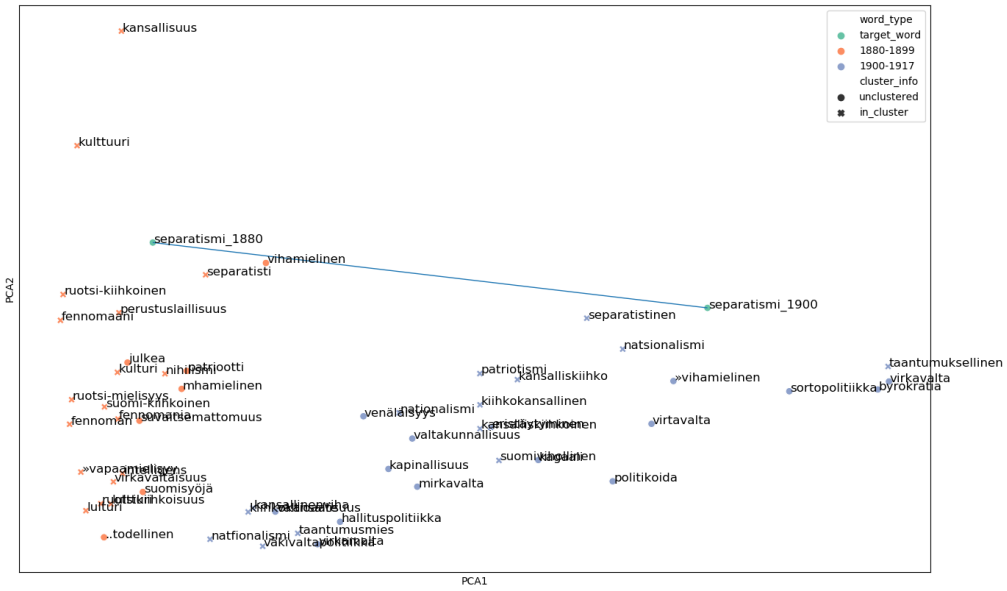
Table 5: Swedish clusters containing word *separatism*

1880-1899	1900-1917
separatismi ‘separatism’ *ruotsi-kiihkoinen* ‘Svekoman’ *ruotsinmielinen* ‘Swedish-minded’	*separatismi* ‘separatism’
ruotsalaisuus ‘Swedishness’ *viikinki* ‘Viking’ *ruotsi-mielinen* ‘Swedish-minded’	*nationalismi* ‘nationalism’ *natsionalismi* ‘nationalism’
fennomaani ‘Fennoman’ *epäkansallinen* ‘anti-national’ *viikingit* ‘Vikings’	*opportunismi* ‘opportunism’ *natfionalismi* ‘nationalism’$_{ocr}$
separatisti ‘separatist’ *ruotsikko* ‘Swedish-minded’(person) *miikinki* ‘Viking’$_{ocr}$ *pöppö* ‘?’	*eristäytyminen* ‘isolation’ *kansalliskiihko* ‘nationalism’
miikingit ‘Vikings’$_{ocr}$ *suomimielinen* ‘Finnish-minded’ *ruotsi-mielisyys* ‘Swedish-mindedness’	*intelligens* ‘intelligence’ *länsieurooppalainen* ‘Western-European’
wiitinki ‘Viking’$_{ocr}$ *wiilinki* ‘Viking’$_{ocr}$ *miitinki* ‘Viking’$_{ocr}$ *ruotsimielinen* ‘Swedish-minded’	*rotutaistelu* ‘race struggle’ *vapaamielisyy* ‘liberalism’$_{ocr}$
suomi-kiihkoinen ‘Fennoman’ *fennoman* ‘Fennoman’ *henkiheimolainen* ‘like minded’	*sanomalehdistö!* ‘press’ *antipatia* ‘antipathy’
dagbladilainen ‘member of the Dagblad circle’ *miiking* ‘Viking’$_{ocr}$ *fennomani* ‘Fennoman’	*kansallinenviha* ‘national anger’ *kiihkokansallisuus* ‘national fervour’
wiiking ‘Viking’$_{ocr}$ *fennomaaninen* ‘Fennoman’ *ruotsikiihkoisuus* ‘Svekomania’	*eristäytyä* ‘self-isolate’ *liittolaisuus* ‘alliance’
wiilinli ‘Viking’$_{ocr}$ *miikinkilehti* ‘Vikings’ newspaper’$_{ocr}$ *suomenmielinen* ‘Finnish-minded’$_{ocr}$	*vihamieli-syy* ‘hostility’$_{ocr}$ *kansallinenylpeys* ‘national pride’
miikinkiläinen ‘Viking (adjective)’$_{ocr}$ *ruolsinmielinen* ‘Swedish-minded’ *ruotsiliihloinen* ‘Svekoman’$_{ocr}$	*kielipolitiikka* ‘language policy’
herranenluokka ‘class of the lords’ *miikingilehti* ‘Vikings’ newspaper’$_{ocr}$ *epälansallinen* ‘anti-national’$_{ocr}$	*kansallinenliike* ‘national movement’

Table 6: Finnish clusters containing word *separatismi*



SWEDISH



FINNISH

Figure 6: PCA plots of *separatism(i)* and its nearest neighbours across time slices. Words marked by $\times$ are part of the separatism cluster in their respective time slice.

DISCUSSION AND FUTURE WORK

The starting point for our inquiry was the assumption that isms became a standard feature in Finnish political, social and cultural language in the course of the nineteenth century. Based on our analysis of newspapers published in Finland since the early nineteenth century to the early twentieth century, this is certainly the case, but the development was somewhat uneven between the two languages. For Swedish, the process was more gradual and more diverse, with a larger selection of isms in use even in the period when the Finnish-language data set is larger. Finnish-

language isms were also surprisingly political compared to those in Swedish, as the language of isms was dominated by discourses of socialism and less by cultural and scientific themes. Here, the Finnish-language clusters show greater continuity than the Swedish-language counterparts.

Our two experiments, clustering isms with one another and clustering individual isms with distributionally similar words (enriched clustering), tell us different things about how the language of isms expanded. The first experiment shows how different isms relate to one another and indicates how the sphere of politics and ideology comes across as a separate category for both languages. Even if medical and artistic isms are associated with political isms through their shared suffix, in language use these domains of life did not intersect much, especially not in Finnish.

The enriched clusters tell us more about the shifts in the distribution and/or meanings of individual isms. In the case of separatism, we can show how the term occupied slightly different domains of discourse in three consecutive time slices. Methodologically, however, we found it important to use textual examples alongside the clusters as clustering and concrete examples provide different views of the conceptual change at hand.

Embeddings and semantics

As we have shown in this paper, the comparison of word embeddings trained on various time periods is a fruitful method for analyzing historical newspapers. Diachronic analysis using vector models is a rapidly growing research field in computational linguistics (see, for example, recent surveys of this topic [Kutuzov et al., 2018, Tahmasebi et al., 2018]).

One research direction is aimed at continuous time representation [Dubossarsky et al., 2019, Gillani and Levy, 2019, Rosenfeld and Erk, 2018, Yao et al., 2018]. These methods reveal gradual semantic changes over time and do not require the data to be divided into discrete time slices.

The most recent approach involves contextual word embeddings, which produce a separate vector for each word mentioned based on its context. Contextualized embeddings are reviewed in [Ethayarajh, 2019] and exemplified by BERT [Devlin et al., 2019] and ELMo [Peters et al., 2018]. These models make it possible to trace differences in word usage across time, though as far as we are aware these models were applied to trace the evolution of a single word—e.g., [Martinc et al., 2020a,b]—rather than detecting the evolution of groups of semantically related words.

Finally, there has been a lot of effort directed towards the development of cross-lingual embeddings [Ruder et al., 2019], which put words from two or more languages into the same vector space and thus enable direct comparison of data from various languages. We suggest that using any of these approaches—namely, contextual, continuous and cross-lingual embeddings—or a combination thereof, might be a productive next step, which would allow for a deeper understanding of the historical development of complex political notions. Using these methods, however, requires statistical evaluation of the output of historical data.

Digital humanities and the study of political vocabularies

The analysis of the history of political thought is not tied to the newest advances in natural language processing, but analyses drawing on them often create space for new interpretations in studying the political imaginaries of past people. In this study of isms as nodes of everyday political thinking in nineteenth-century newspapers from Finland, we have produced new and reliable ways of charting and visualizing the expansion of the vocabulary of isms. Our method is particularly noteworthy that it can grasp developments in word use that relate both to growth

in frequency and changes in the distribution of the word. Thus our findings regarding the importance of socialism as a political keyword are not surprising to someone with good knowledge of political vocabulary in Finland, but our method shows the sheer amounts and pivotal role of socialism in a way that has not previously been possible. Nor have there been any attempts to compare the discourse of socialism across the language divide in Finland. The findings relating to separatism are different in the sense that we were not expecting to find anything out of the ordinary relating to it. We were rather surprised that it emerged as a interesting case based on a semi data-driven perspective.

Our cases relating to socialism and separatism also indicate that the relationship between distribution and meaning, as pointed out in the so-called distributional hypothesis, which is usually attributed to Zellig Harris [Harris, 1970, Sahlgren, 2008], is not as straightforward as sometimes believed.[14] While there is a link between the change in distribution and semantic change, this link seems to be easier to capture in clear cases of polysemy than in relatively vague and flexible terms such as the isms studied here. Isms are often also in hierarchical relation with one another, especially when qualified in some way. For instance, the words state socialism (*statssocialism*) and municipal socialism (*kommunalsocialism*) are found in Table 8. The former clusters together with socialism but not the latter. This suggests that the clustering is related more to social meaning than to strict semantic meaning.

While word embeddings and other methods of analyzing the distribution of terminology are increasingly looking for new avenues in studying multilingual corpora, we wish to further point out that the case of isms may be a fruitful avenue for developing multilingual approaches. Dealing with Finnish and Swedish in one country showed that the historical translatability between the languages (even if Finnish is less prone to introduce new isms) can be very useful in studying political vocabularies and thinking in different linguistic contexts. While a comparison cross state borders requires good contextual knowledge that takes into account both linguistic and political specificities, the fact that historical actors readily translated isms as cognates is an exceptionally good starting point for cross-lingual analysis.

ACKNOWLEDGEMENTS

We are grateful to Simon Hengchen and Mark Granroth-Wilding for the help with data preparation. This work has been supported by the European Union Horizon 2020 research and innovation program under grants 770299 (NewsEye) and 825153 (EMBEDDIA).

References

Domagoj Alagić, Jan Šnajder, and Sebastian Padó. Leveraging lexical substitutes for unsupervised word sense induction. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Risto Alapuro, Ilkka Liikanen, Kerstin Smeds, and Henrik Stenius, editors. *Kansa liikkeessä*. Kirjayhtymä, Helsinki, 1987.

Duncan Bell. What Is Liberalism? *Political Theory*, 42(6):682–715, December 2014. ISSN 0090-5917, 1552-7476. doi: 10.1177/0090591714535103.

Cesare Cuttica. To use or not to use ... the intellectual historian and the isms: A survey and a proposal. *Études Épistémè*, 23, 2013. ISSN 1634-0450. doi: 10.4000/episteme.268.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. pages 4171–4186, 2019.

Haim Dubossarsky, Simon Hengchen, Nina Tahmasebi, and Dominik Schlechtweg. Time-out: Temporal referencing for robust modeling of lexical semantic change. In *The 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019.

[14] We thank Antti Kanner for pointing this interpretation out to us.

Max Engman. *Språkfrågan: Finlandssvenskhetens uppkomst 1812–1922*. Svenska litteratursällskapet i Finland, 2016. ISBN 978-951-583-354-9.

Kawin Ethayarajh. How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1006.

Manaal Faruqui, Yulia Tsvetkov, Pushpendre Rastogi, and Chris Dyer. Problems with evaluation of word embeddings using word similarity tasks. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*, pages 30–35, 2016.

Michael Freeden, Javier Fernández-Sebastián, and Jörn Leonhard. *In search of European liberalisms: Concepts, languages, ideologies*. Berghahn Books, New York, 2019. ISBN 978-1-78920-280-9.

Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814): 972–976, 2007.

Nabeel Gillani and Roger Levy. Simple dynamic word embeddings for mapping perceptions in the public sphere. In *NAACL HLT 2019*, page 94, 2019.

Istvan Hahn. Die Begriffe auf â“ismos. In C. Welskopf, editor, *Soziale Typenbegriffe im alten Griechland und ihr Fortleben in den Sprachen der Welt: Band 4, Untersuchungen ausgewählter altgriechischer sozialer Typenbegriffe und ihr Fortleben in Antike und Mittelalter*, pages 52–99. Akademie-Verlag, Berlin, 1981.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2016, page 2116. NIH Public Access, 2016.

Zellig Harris. Distributional structure. *Papers in structural and transformational Linguistics*, pages 775–794, 1970.

Simon Hengchen, Ruben Ros, and Jani Marjanen. A data-driven approach to the changing vocabulary of the 'nation' in English, Dutch, Swedish and Finnish newspapers, 1750–1950. In *In Proceedings of the Digital Humanities (DH) conference 2019, Utrecht, The Netherlands*, 2019.

H. M. Höpfl. Isms. *British Journal of Political Science*, 13(1):1–17, 1983. ISSN 1469-2112, 0007-1234. doi: 10.1017/S0007123400003112.

Matti Hyvärinen, Jussi Kurunmäki, Kari Palonen, Tuija Pulkkinen, and Henrik Stenius, editors. *Käsitteet liikkeessä: Suomen poliittisen kulttuurin käsitehistoria*. Vastapaino, Tampere, 2003. ISBN 978-951-768-130-8.

Helge Jordheim. Introduction: Multiple times and the work of synchronization. *History and Theory*, 53(4):498–518, 2014.

Helge Jordheim. Synchronizing the world: Synchronism as historiographical practice, then and now. *History of the Present*, 7(1):59–95, 2017.

Osmo Jussila. *Suomen suuriruhtinaskunta 1809–1917*. WSOY, Helsinki, 2004. ISBN 978-951-0-29500-7.

Mikko Kemppainen. *Sosialismin, uskonnon ja sukupuolen dynamiikkaa: 1900-luvun alun työväenliikkeen naiskirjailijat aatteen määrittelijöinä*. Työväen historian ja perinteen tutkimuksen seura, Tampere, 2020.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. Temporal analysis of language through neural language models. *ACL 2014*, page 61, 2014.

Jussi Kurunmäki and Jani Marjanen. Isms, ideologies and setting the agenda for public debate. *Journal of Political Ideologies*, 23(3):256–282, 2018a. doi: 10.1080/13569317.2018.1502941.

Jussi Kurunmäki and Jani Marjanen. A rhetorical view of isms: an introduction. *Journal of Political Ideologies*, 23(3):241–255, 2018b. ISSN 1356-9317, 1469-9613. doi: 10.1080/13569317.2018.1502939.

Jussi Kurunmäki and Ilkka Liikanen. The Formation of the Finnish Polity within the Russian Empire: Language, Representation, and the Construction of Popular Political Platforms, 1863–1906. *Harvard Ukrainian Studies*, 35(1-4):399–416, 2018.

Andrey Kutuzov, Elizaveta Kuzmenko, and Lidia Pivovarova. Clustering of Russian adjective-noun constructions using word embeddings. In *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, pages 3–13, 2017.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397, 2018.

Geoffrey N. Leech. *Semantics*. Penguin, Harmondsworth, 1974. ISBN 978-0-14-021694-3.

Jörn Leonhard. *Liberalismus: Zur historischen Semantik eines europäischen Deutungsmusters*. Veröffentlichungen des Deutschen Historischen Instituts London. R. Oldenbourg, München, 2001.

Fang Li and Xiaojie Wang. Improving word embeddings for low frequency words by pseudo contexts. In *Chinese*

Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, pages 37–47. Springer, 2017.

Eetu Mäkelä. Las: an integrated language analysis tool for multiple languages. *The Journal of Open Source Software*, 1, 2016.

Jani Marjanen. "Svenskar äro vi icke mera": Om ett uttrycks historia. In Henrika Tandefelt, Julia Dahlberg, Aapo Roselius, and Oula Silvennoinen, editors, *Köpa salt i Cádiz och andra berättelser*. Siltala, Helsinki, 2020.

Jani Marjanen, Ville Vaara, Antti Kanner, Hege Roivainen, Eetu Mäkelä, Leo Lahti, and Mikko Tolonen. A National Public Sphere? Analyzing the Language, Location, and Form of Newspapers in Finland, 1771–1917. *Journal of European Periodical Studies*, 4(1):54–77, June 2019. ISSN 2506-6587. doi: 10.21825/jeps.v4i1.10483.

Matej Martinc, Syrielle Montariol, Elaine Zosa, and Lidia Pivovarova. Capturing evolution in word usage: Just add more clusters? In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan*, 2020a.

Matej Martinc, Petra Kralj Novak, and Senja Pollak. Leveraging contextual embeddings for detecting diachronic semantic shift. In *LREC*, 2020b.

Wiktor Marzec and Risto Turunen. Socialisms in the Tsarist Borderlands. *Contributions to the History of Concepts*, 13(1):22–50, June 2018. ISSN 1807-9326, 1874-656X. doi: 10.3167/choc.2018.130103.

Tomas Mikolov, Kai Chen, Greg S Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *NIPS*, 2013.

Yrjö Nurmio. *Suomen sensuuriolot Venäjän vallan alkuaikoina vv. 1809–1829*. WSOY, Helsinki, 1934.

Yrjö Nurmio. *Taistelu suomen kielen asemasta 1800-luvun puolivälissä: Vuoden 1850 kielisäännöksen syntyhistorian, voimassaolon ja kumoamisen selvittelyä*. WSOY, Helsinki, 1947.

Tuula Pääkkönen, Jukka Kervinen, Asko Nivala, Kimmo Kettunen, and Eetu Mäkelä. Exporting Finnish digitized historical newspaper contents for offline use. *D-Lib Magazine*, 22(7/8), 2016.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, 2018.

Heli Rantala, Hannu Salmi, Aleksi Vesanto, and Filip Ginter. Tekstien pitkä elämä: Ajassa liikkuvat tekstit suomalaisessa sanomalehdistössä 1771–1920. *Ennen ja Nyt*, (2), 2019. URL https://www.ennenjanyt.net/2019/08/tekstien-pitka-elama-ajassa-liikkuvat-tekstit-suomalaisessa-sanomalehdistossa-1771-19

Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA.

kirjoittaja Rosenblatt, Helena. *The lost history of liberalism: From ancient Rome to the twenty-first century*. Princeton University Press, Princeton, 2018.

Alex Rosenfeld and Katrin Erk. Deep neural models of semantic shift. In *NAACL HLT 2018*, pages 474–484, 2018.

Sebastian Ruder, Ivan Vulić, and Anders Søgaard. A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*, 65:569–631, 2019.

Magnus Sahlgren. The distributional hypothesis. *Italian Journal of Linguistics*, 20:33–53, 2008.

Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. A wind of change: Detecting and evaluating lexical semantic change across times and domains. *arXiv preprint arXiv:1906.02979*, 2019.

Tobias Schnabel, Igor Labutov, David Mimno, and Thorsten Joachims. Evaluation methods for unsupervised word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 298–307, 2015.

Philippa Shoemark, Farhana Ferdousi Liza, Dong Nguyen, Scott Hale, and Barbara McGillivray. Room to glo: A systematic comparison of semantic change detection approaches with word embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 66–76, 2019.

Ivo Spira. *A conceptual history of Chinese -isms: The modernization of ideological discourse, 1895–1925*. Number Volume 4 in Conceptual history and Chinese linguistics. Brill, 2015. ISBN 978-90-04-28787-7.

Henrik Stenius. The Finnish citizen: How a translation emasculated the concept. *Redescriptions: Political Thought, Conceptual History and Feminist Theory*, 8:172–188, 2004.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. Survey of computational approaches to diachronic conceptual change. *arXiv preprint arXiv:1811.06278*, 2018.

Päiviö Tommila and Raimo Salokangas. *Sanomia kaikille: Suomen lehdistön historia*. Kleio ja nykypäivä. Edita, Helsinki, 1998. ISBN 978-951-37-2621-8.

Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. Dynamic word embeddings for evolving semantic discovery. In *The 11th ACM Conference on Web Search and Data Mining*, 2018.

Clas Zilliacus and Henrik Knif. Opinionens tryck: En studie över pressens bildningsskede i Finland, 1985. ISBN: 9789519018119 Place: Helsingfors Series: Skrifter utgivna av Svenska litteratursällskapet i Finland, nr. 526.

ANNEX 1: ISM CLUSTERS FOR FINNISH DATA

Table 7: Clustering obtained for Finnish words ending with the ism suffix. We show cluster words and their frequencies in the respective time slice, sorted by frequency. Cluster *representatives* are marked in italics.

1860-1879

sosialismi	172

1880-1899

sosialismi	5115
anarkismi	1120
nihilismi	602
militarismi	328
kommunismi	316
parlamentarismi	312
kapitalismi	301
liberalismi	236
radikalismi	171
boulangismi	163
kosmopolitismi	128
protestantismi	115
baptismi	108

realismi	1029
pessimismi	614
idealismi	351
symbolismi	291
naturalismi	279
optimismi	182

separatismi	924
patriotismi	231
fanatismi	126

organismi	415
magnetismi	328

pietismi	370
materialismi	367
ateismi	167
metodismi	147
dualismi	105
despotismi	101

hypnotismi	733
spiritismi	542
alkoholismi	527

reumatismi	1706

teismi	119

1900-1917

realismi	2097
pessimismi	1192
idealismi	591
naturalismi	794
pietismi	476
impressionismi	342
aforismi	262
humanismi	242
symbolismi	231
panteismi	230
egoismi	201
kubismi	169
asketismi	154
fatalismi	152
altruismi	150
mystisismi	141
klassisismi	123
ratsionalismi	119

materialismi	3232
spiritismi	1327
hypnotismi	623
monismi	449
darwinismi	435
modernismi	174
marxismi	148
pragmatismi	113

organismi	1009
magnetismi	433
mikro-organismi	130

mekanismi	564

sosialismi	75117
anarkismi	5630
terrorismi	2063
kommunismi	1548
sialismi	910
syndikalismi	524
individualismi	398
nihilismi	397
ateismi	341
antimilitarismi	306
revisionismi	288
sofalismi	178
remisionismi	139
indimidualismi	127
rialismi	117
darvinismi	111
antisemitismi	102

>sosialismi	832
..sosialismi	316
sionismi	221
vegetarismi	196
”sosialismi	170
.sosialismi	163
sosialismi	133
’sosialismi	126
sosialismi	108

reumatismi	9629
matismi	180
nivelreumatismi	158

militarismi	7062
imperialismi	2796
despotismi	661
absolutismi	350
tsarismi	318
huliganismi	270
panslavismi	202
vandalismi	194
bolshevismi	194
hellenismi	187
klerikalismi	147
klerkalismi	146
germanismi	104

separatismi	2008
natsionalismi	1852
optimismi	1580
patriotismi	993
nationalismi	657
fanatismi	589
nalismi	134
anakronismi	129
natfionalismi	120

fotsialismi	203
fofalismi	151
anarfismi	126
fofialismi	123

onnismi	517
onanismi	265

kapitalismi	20681
talismi	388
industrialismi	346
suurkapitalismi	302
barbarismi	225
tpitalismi	178
lpitalismi	166
tapitalismi	155
pitalismi	137
lapitalismi	113
kapitalismi	104

parlamentarismi	3413
liberalismi	1156
radikalismi	645
feodalismi	438
opportunismi	420
dualismi	293
valtiososialismi	235
protestantismi	213
gmerkantilismi	140

alkoholismi	4312
kunnallinensosialismi	1085
alloholismi	105
holismi	158

teismi	598
tarismi	175
turismi	126

reumatismi	212
.reumatismi	122

ANNEX 2

Table 8: Clustering obtained for Swedish words ending with ism suffix. We show cluster words and their frequencies, sorted by frequency. Cluster *representatives* are marked in italics.

1820-1839

patriotism	232
fanatism	165
despotism	153

1840-1859

patriotism	581
egoism	560
fanatism	431
socialism	294
pauperism	290
despotism	254
kommunism	160
liberalism	136
radikalism	105

organism	410
mekanism	217
magnetism	170
galvanism	124

rheumatism	101

pietism	505
protestantism	342
katholicism	155

1860-1879

patriotism	2664
liberalism	1148
materialism	461
konservatism	446
dualism	347
parlamentarism	341
absolutism	265
optimism	228
pessimism	209
rationalism	158
konstitutionalism	139
anakronism	137
protektionism	135
sofism	100

rheumatism	571
reumatism	305
galvanism	151

socialism	1263
katolicism	988
protestantism	846
kommunism	363
nihilism	282
katholicism	272
mormonism	240
pauperism	211
skandinavism	211
jesuitism	207
spiritism	188
panslavism	163
darwinism	112
ateism	102

baptism	360
pietism	205
separatism	118

despotism	923
radikalism	585
ultramontanism	458
bonapartism	337
klerikalism	177
imperialism	150
carlism	141
federalism	136
republikanism	115

fanatism	1710
terrorism	325
vandalism	233
cynism	190
chauvinism	115

antagonism	336
schism	323

egoism	1024
realism	388
idealism	183
heroism	153
mysticism	142
naturalism	115
dilettantism	106

organism	1575
mekanism	992
magnetism	328
statsorganism	120

Table 8: Clustering obtained for Swedish words ending with the ism suffix: continuation

1880-1899

socialism	5560
katolicism	2154
anarkism	1975
protestantism	1408
militarism	1366
nihilism	1181
antisemitism	911
absolutism	641
statssocialism	485
kommunism	459
journalism	244
bimetallism	212
jesuitism	206
nationalism	198
individualism	196
utilitarism	124
germanism	115

baptism	641
mormonism	503
sekterism	366
metodism	259
finlandism	223
laestadianism	132
fennicism	106

separatism	829
partifanatism	278
språkfanatism	273
nepotism	121

vandalism	572
anakronism	298

patriotism	4792
liberalism	2705
konservatism	1806
parlamentarism	1688
radikalism	1455
protektionism	1222
chauvinism	950
despotism	868
opportunism	344
skandinavism	311
konstitutionalism	259
republikanism	203
feodalism	101

mekanism	3237
hypnotism	1811
magnetism	932
idiotism	287
jordmagnetism	175
galvanism	169
somnambulism	138
bypnotism	132
atavism	106

schism	1263
antagonism	863
dualism	467
statsorganism	146

aforism	283
darwinism	276
darvinism	165
vegetarianism	154

egoism	3057
materialism	1003
pietism	547
formalism	482
ateism	455
rationalism	276
obskurantism	221
positivism	221
indifferentism	213
industrialism	136
asketism	131
barbarism	123

realism	2295
naturalism	1134
idealism	834
symbolism	561
mysticism	422
dilettantism	309
sofism	309
humanism	216
kosmopolitism	171

spiritism	1123
kannibalism	383
buddism	175
muhamedanism	166
buddhaism	151
spiritualism	103

alkoholism	1364
pauperism	231
morfinism	129

boulangism	1128
terrorism	707
klerikalism	569
panslavism	341
kapitalism	226
hellenism	206
partikularism	180
imperialism	151
bonapartism	143
ultramontanism	129
kollektivism	121
cesarism	110

fanatism	3086
pessimism	1382
cynism	846
optimism	839
skepticism	548
heroism	320
fatalism	310
lokalpatriotism	112

reumatism	5735
ledgångsreumatism	1381
rheumatism	1262
matism	274
ledgångsrheumatism	188
ledgängsrheumatism	126

organism	5713
mikroorganism	621
djurorganism	110

amerikanism	161

Table 8: Clustering obtained for Swedish words ending with ism suffix: continuation

1900-1917

socialism	15080
parlamentarism	2231
liberalism	2034
konservatism	2034
imperialism	1637
radikalism	1438
absolutism	879
klerikalism	818
konstitutionalism	695
protektionism	650
skandinavism	440
opportunism	288
marxism	286
internationalism	285
proportionalism	245
demokratism	234
statssocialism	204
aktivism	168
oktobrism	136
försvarsnihilism	117
monarkism	112
modernism	109

alkoholism	2829
vegetarism	245
darwinism	193
kommunalsocialism	145
vegetarianism	130
nykterism	112

antagonism	943
dualism	415
statsorganism	177
parallellism	105

schism	3237
skism	167

turism	448

idealism	1113
materialism	694
individualism	512
spiritism	506
pietism	415
mysticism	266
sofism	260
journalism	189
humanism	179
indifferentism	178
kosmopolitism	167
rationalism	158
obskurantism	156
ateism	146
asketism	138
atavism	129
dogmatism	121
monism	114

realism	1785
naturalism	560
impressionism	319
symbolism	247
dilettantism	245
kubism	225
klassicism	183

slavism	317
germanism	240
panslavism	211
hellenism	141
pangermanism	130

reumatism	6423
rheumatism	820
muskelreumatism	142

ledgångsreumatism	1811
matism	470

anarkism	1709
terrorism	1600
syndikalism	387
antisemitism	341
antimilitarism	267
kommunism	256
feminism	242
jesuitism	180
revisionism	166
nihilism	125
ungsocialism	112
kollektivism	110

nationalism	5398
patriotism	4254
separatism	1079
chauvinism	1002
språkfanatism	373
suometarianism	363
fariseism	134

katolicism	1363
protestantism	726
kannibalism	338
buddism	261
buddhism	154
muhammedanism	150

organism	6627
mekanism	2332
mikroorganism	453
samhällsorganism	125

optimism	2565
pessimism	2023
skepticism	563

aforism	396
finlandism	221

egoism	2942
fanatism	2496
cynism	900
heroism	482
fatalism	252
partifanatism	219
lokalpatriotism	172
altruism	145
klassegoism	106
knutpatriotism	102

kapitalism	2399
militarism	2346
despotism	917
industrialism	732
tsarism	568
barbarism	169
utilitarism	142
feodalism	132
storkapitalism	128

vandalism	703
byråkratism	426
formalism	496
anakronism	423
nepotism	117
servilism	106

hypnotism	528
magnetism	362
idiotism	133
jordmagnetism	116

baptism	207
mormonism	125
sekterism	121

polism	182